

首位数规律在物理学中的应用

邵立晶, 马伯强

北京大学物理学院;核物理与核技术国家重点实验室,北京 100871

摘要 自然世界数据库中的首位非零数字并不以 1~9 的等概率出现,而是符合一个对数分布,即 Benford 定理。在研究该定理在粒子物理与天体物理中的表现时发现,粒子物理中的强子宽度与寿命、脉冲星中的功率与运动等多类物理量均符合 Benford 分布;统计物理中的三大正则统计——玻尔兹曼-吉布斯统计、费米-狄拉克统计和玻色-爱因斯坦统计均解析满足首位数规律。本文简介了首位数定理,阐述了其在粒子物理、脉冲星中的应用,讨论其标度不变性、整数幂次不变性、基数不变性等性质,最后分析了三大物理统计中的首位数规律。如何更好地理解 Benford 定理这个“百年难题”,探讨首位数分布的自然规律,还需要更多的后续研究。

关键词 统计物理;首位数;强子宽度;脉冲星

中图分类号 N49

文献标识码 A

文章编号 1000-7857(2010)18-0098-05

Applications of the First Digit Law in Physics

SHAO Lijing, MA Boqiang

State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing 100871, China

Abstract The occurrence of the nonzero leftmost digit, i.e., 1, 2, ..., 9, of numbers from many real world sources is not uniformly distributed as one might naively expect, but instead, the nature favors smaller ones according to a logarithmic distribution, named Benford's law. This paper discusses systematically the first digit distributions of physical quantities in the fields of particle physics and astrophysics, and it is found that the full widths of hadrons and various quantities of pulsars conform to Benford's law very well. The first and second digit distributions of derivatives of period and frequency of pulsars are also found to obey the n -digit Benford's law. It is further pointed out that the pulsar data can serve as an ideal assemblage to study the first digit distributions of the real world data. Furthermore, three kinds of widely used physical statistics, i.e., the Boltzmann-Gibbs distribution, the Fermi-Dirac distribution, and the Bose-Einstein distribution, are found to conform to this law in an analytical manner. Thus, a new physical perspective is used to look into the underlying reason of this law. Moreover, various elegant mathematical properties of Benford's law are discussed in detail.

Keywords physical statistics; first digit distribution; hadron full widths; pulsars

1 首位数定理简介

首位数指的是数字的首位非零数位,其取值范围为 1~9 的整数,如 314、0.0314、31400 的首位数均为 3。然而首位数从 1 到 9 的取值概率并不相等,自然界的首位数规律满足一个对数分布,即 Benford 定理,也称首位数定理^[1-3]。

首先发现首位数规律的是 Newcomb,他发现对数表前面的页码脏得比后面的快。Newcomb 指出这个现象背后的原因是,在自然生活中,人们所使用数字的指数尾数的对数是均

匀分布的^[3],但当时他的理论并没有引起科学界的重视。

1938 年,美国通用电气公司的 Benford 也注意到了对数表的页码变脏的不均匀性的现象,并提出了首位数定理^[1]。科学界对该定理产生了浓厚兴趣,并将其命名为 Benford 定理。

自首位数定理被发现的 100 多年来,来自物理学、数学、经济学等不同领域的专家,从不同的认知角度,对它进行了深入研究与探讨,取得了不少重大进展。然而,其深刻原因至

收稿日期:2010-05-04;修回日期:2010-08-27

基金项目:国家自然科学基金委员会创新研究群体科学基金项目(10721063);国家自然科学基金面上项目(10975003);国家基础科学人才培养基金项目(J0630311、J0730316)

作者简介:邵立晶,研究方向为 Benford 定理、强子结构模型、洛伦兹破坏唯像学等,电子信箱:lshao@pku.edu.cn;马伯强(通信作者),教授,研究方向为粒子物理、中高能核物理、高能宇宙线物理与宇宙学、数学物理方法在其他学科中的应用等,电子信箱:mabq@pku.edu.cn

今未能彻底理解。

自然界数据的第一个非零数字出现 1~9 的概率并不相等,而是符合一个对数公式,其中首位数字为 k 的概率为^[1]

$$P(k) = \lg(1 + 1/k) \quad (k=1, 2, \dots, 9) \quad (1)$$

式(1)即为首位数定理。首位数偏向于出现较小的数字,其具体数值参见表 1。

表 1 Benford 定理预言的首位数为 1~9 出现的概率

Table 1 Occurrence of the first digit from 1 to 9 according to Benford's law.

首位数	1	2	3	4	5	6	7	8	9
Benford 定理	0.301	0.176	0.125	0.097	0.079	0.067	0.060	0.051	0.046

人们在很多毫无关联的自然数据库中都可以看到首位数定理的表现,如街道住址、原子分子的重量、湖泊的面积、河流的长度^[1]、物理学常数^[4]、 α 衰变的半衰期^[5-7]、存活率^[8]、复杂原子跃迁光谱中的电偶极线强度^[9]、eBay 的交易额^[10]、股票指数^[11]、个人计算机中的文件大小^[2]、简单的动力学系统^[12-14]等。其中,南京大学的任中洲小组也在 α 衰变半周期的首位数分布方面做出了相应的贡献^[6-7]。然而,也存在一些不符合 Benford 定理的分布。但至今无法先验地判断某个数据库是否符合 Benford 定理,只能仔细研究和计数后才能得出结论。

首位数定理已在社会学与自然科学领域有了广泛的应用。若某个数据集已被验证为符合 Benford 规律,那么可通过检验首位数对同类实际数据的合理性进行检验。这种方法已被应用于检验会计与财务里的伪造情况^[15-17]、选举中的欺骗事例^[2]、科学实验数据的造假行为^[18]等。另外,由于绝大多数的自然界数据符合这个规律,如果在计算机的算法与内存的设计中考虑到这个因素,可有效地提高计算机的运行速度^[9]、减小内存消耗^[20]、优化图像处理^[21-22]等。

2 首位数定理在粒子物理中的应用

首位数定理在自然界的粒子物理领域也是成立的。北京大学马伯强课题组系统地研究了粒子物理中的首位数定理,并首次发现介子、重子的首位数分布非常满足 Benford 定理^[23],微观领域中的强子(介子与重子的统称)宽度遵守 Benford 分布^[24]。在统计研究中,尽量使用多种方式无偏见地处理强子数据,并系统地使用 Pearson 卡方进行拟合优度的检验:

$$\chi^2(n-1) = \sum_{i=1}^n \frac{(N_{Obs} - N_{Ben})^2}{N_{Ben}}$$

其中, N_{Obs} 为观测计数, N_{Ben} 为 Benford 定理理论预言值, $n-1$ 为卡方自由度,在粒子物理中的检验中卡方自由度为 8,在 95% 的置信水平下的临界卡方为 $\chi^2(8) = 15.507$ 。多种情况下的检验结果非常理想,可以接受“介子、重子和强子的宽度符合 Benford 分布”的零假设。介子、重子和强子的首位数分布参见图 1^[24],图中的介子、重子和强子的检验卡方分别为 6.32、4.57、4.62; 其中, Benford 理论预言的允许误差来自于二项式误差估算 $[N_{Ben} \cdot P(k) \cdot (1-P(k))]^{1/2}$ 。

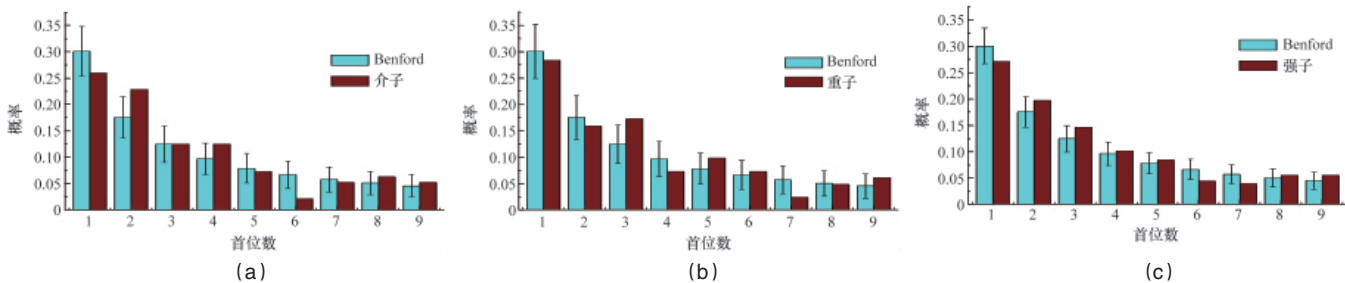


图 1 介子(a)、重子(b)和强子(c)宽度的首位数分布与 Benford 定理的比较

Fig. 1 Comparisons between the first digit distributions of full widths of mesons (a), baryons (b), and hadrons (c) with Benford's law

3 首位数定理在脉冲星中的应用

Benford 定理在宏观的天体物理领域也成立。马伯强课题组系统地研究了澳大利亚国家望远镜装置脉冲星数据库^[25]中近 2000 个脉冲星数据的首位数分布,发现脉冲星的很多物理量都符合首位数规律,如脉冲星的年龄、周期变化率、频率变化率、自转变慢能损率、到日点能量流量、400 和 1400MHz 射电波段光度和流量强度、经度和纬度方向的自行运动速度等^[26]。部分物理量虽然也显示了数位出现概率随着数位变大而递减的趋势,但 Pearson 卡方统计检验却拒绝了“符合

Benford 分布”这一零假设,如脉冲星的周期和频率、脉冲星伴星质量的上限和下限等。其中,脉冲星的伴星质量的上限和下限不满足首位数分布规律,是因为它们并非是客观的物理观测,强加的关于倾斜角等的假设和实际的物理约束限制它们成为很好的 Benford 数据集^[26]。而对于周期和频率,满足推广的 Benford 定理。图 2(a)为脉冲星周期的首位数分布和 Benford 定理理论预言的比较,2(b)为其与推广的 Benford 定理(gBenford)的对比,拟合上有了明显的改进,也与严格的 F 检验结论一致^[26]。

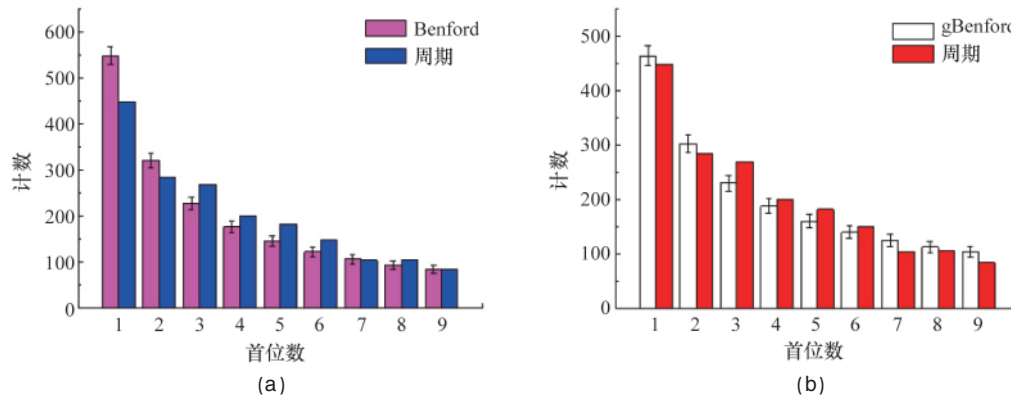


图2 脉冲星周期的首位数分布与 Benford 定理(a)和推广的 Benford 定理(b)的比较
Fig. 2 Comparisons between the first digit distribution of periods of pulsars with Benford's law (a) and the generalized Benford's law (b)

系统研究指出,由于脉冲星数据库容量大、数据类型多,可作为一个天然的检验与研究首位数定理的理想数据库。另外,可通过比较脉冲星模型的首位数分布与 Benford 分布,进而对脉冲星理论做出约束和预言^[26]。

4 首位数定理若干性质的讨论

Benford 定理之所以能引起科学界的关注,是因为它有深刻的数学物理性质。

4.1 首 n 位数的 Benford 定理

Benford 定理预言,前 n 位数分别为 d_1, d_2, \dots, d_n 的概率为

$$P_{\text{Ben}}(d_1, d_2, \dots, d_n) = \lg \left[1 + \left(\sum_{i=1}^n 10^{n-i} d_i \right)^{-1} \right]$$

如,前两位数分别为 1 和 5 的概率是 $\lg(1+1/15)$ 。脉冲星的周期与频率对时间的导数的首两位数分布与首 n 位 Benford 定理的比较参见图 3,符合较好, Pearson 卡方检验也给出了较好的拟合优度。

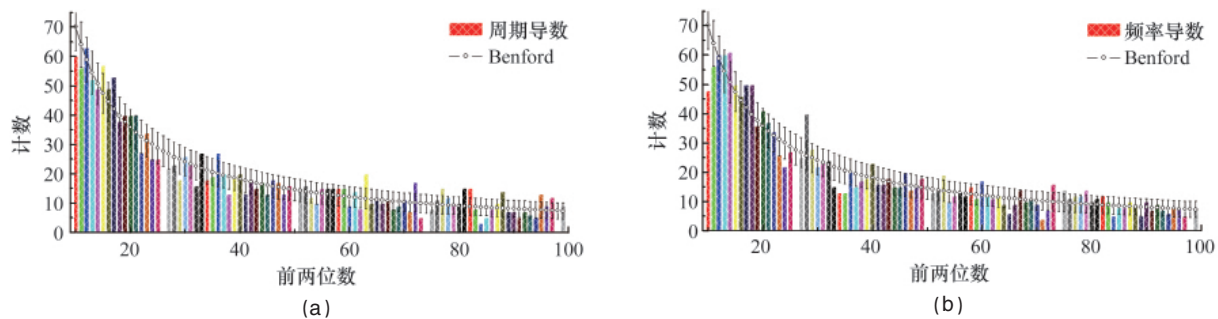


图3 脉冲星的周期(a)与频率(b)对时间的导数的首两位数分布与 Benford 定理的比较
Fig. 3 Comparisons between the first and second digits of derivatives of period (a) and frequency (b) of pulsars with n -digit Benford's law

4.2 对数空间的 Benford 定理

1881 年, Newcomb 指出,首位数规律可以理解为指数尾数的对数的均匀分布^[3],换句话说,若将数据集写成 $\{10^{n+f}\}$ 的形式,其中 n 为整数, f 为小数,那么, f 在 $[0, 1)$ 上的均匀分布即意味着首位数规律。

4.3 标度不变性

1961 年, Pinkham 指出,首位数定理具有标度不变性^[27],即符合 Benford 分布的数据集,乘以任意一个非零常数后仍符合 Benford 分布。实际上,经 Benford 数据集 $\{N\}$ 乘以一个非零常数 C ,由于 $\lg CN = \lg C + \lg N$,根据 Benford 定理在对数空间的性质,很容易就能得到标度不变性。数学上, Benford 定理是唯一具有标度不变性的数位定理^[28]。

物理上,标度不变性意味着首位数定理不依赖于具体的物理单位的选择, Burke 和 Kincaid^[4]关于物理常数在国际单位制与英国单位制下的讨论也可清楚地说明这一点。

4.4 整数幂次不变性

同样,对于任意非零整数 n ,由于 $\lg N^n = n \lg N$, Benford 数据集的非零整数次幂也满足首位数定理^[24]。

根据定义,强子宽度 Γ 与强子寿命 τ 之间存在关系: $\Gamma \tau = \hbar$,其中 \hbar 为约化普朗克常数,则 $\tau = \hbar \Gamma^{-1}$ 。

应用标度不变性和整数幂次不变性可以得到,强子寿命 τ 也满足 Benford 分布^[24]。类似地,脉冲星的周期和频率的首位数分布并不是独立的,这一点在推广的 Benford 定理中能更清楚地看到^[26]。



4.5 基数不变性

首位数分布具有基数不变性^[29-36],也就是说,如果数据集在十进制下是满足首位数规律的,那么将其转换成 d 进制后,满足基数为 d 的 Benford 分布:

$$P(k) = \log_d \left(1 + \frac{1}{k} \right) \quad (k=1, 2, \dots, d-1)$$

Hill 在概率论的框架下,从数学上证明了“标度不变性意味着基数不变性”^[32]和“基数不变性意味着 Benford 分布”^[33]。另外,前面所讨论的前 n 位数的 Benford 分布,其实也可看成是 100 进制(前 2 位数)、1000 进制(前 3 位数)下的结论。

4.6 相乘过程与推广的 Benford 定理

2001 年, Pietronero 等^[37]对 Benford 定理提到了一种新的理解,假设某个量 $m(t)$ 随时间演化的过程为随机相乘的过程,在隔时间 δt 后, $m(t)$ 变为

$$m(t+\delta t) = r(t)m(t) \quad (2)$$

其中, $r(t)$ 为随机变量。这种模式在股票市场等领域很容易认识与理解。对式(2)取对数,有

$$\lg m(t+\delta t) = \lg m(t) + \lg r(t)$$

将 $\lg r(t)$ 看成一个新的随机变量,可见在对数空间,这是一个随机布朗运动过程;当时间足够长后,运用大数定理,可知其方差为无穷大, $\lg m(t)$ 趋于均匀分布,从而有

$$P(k) = \frac{\int_k^{k+1} d \lg m(t)}{\int_1^{10} d \lg m(t)} = \frac{\int_k^{k+1} m^{-1} dm}{\int_1^{10} m^{-1} dm} = \lg \left(1 + \frac{1}{k} \right) \quad (3)$$

式(3)即为 Benford 分布^[37]。另外,从第二个等号可见, m 的概率密度反比于 m ,这是该定理一个很重要的性质^[37-38]。

Pietronero 等^[37]将概率密度推广为正比于 $m^{-\alpha}$, $\alpha=0$ 对应着

首位数的均匀分布, $\alpha=1$ 对应着 Benford 分布。运用推广的 Benford 定理可知,脉冲星的周期与频率分别满足 $\alpha=0.8$ 和 $\alpha=1.2$ 的分布,周期首位数的分布与推广的 Benford 定理的比较如图 2(b)所示^[26]。

然而,推广的 Benford 分布中,指数 α 的物理意义还需进一步研究。

5 首位数定理与三大物理统计

物理学上的三大正则统计——玻尔兹曼-吉布斯统计(BG)、费米-狄拉克统计(FD)和玻色-爱因斯坦统计(BE),其首位数分布均能以某种方式回归到 Benford 分布^[38]。

对于 BG 统计和 FD 统计,其首位数分布轻微地依赖于系统的温度 T ,或者等价地依赖于 $\beta=1/kT$, k 为玻尔兹曼常数。在 β 的对数空间内, BG 分布和 FD 分布的首位数呈现一种以 Benford 分布为中心值的小型周期振荡。其中, BG 分布的首位数分布如图 4^[38]所示(同样可参考文献[39]中的图 1),图中的水平直线为 Benford 分布的理论值。FD 分布的情况类似^[38]。对于 BE 分布,分析发现,不论温度为多少,它都精确地满足首位数定理^[38]。

虽然 BG 分布和 FD 分布对于 Benford 定理存在一种周期性偏离,但其中心值为 Benford 定理的理论值,偏离的振幅也非常小。如果要在 95% 的置信区间下区分这两种物理分布与 Benford 分布, Pearson 卡方检验需要上千个数据^[38],而上千个实际数据的搜寻是不现实的。并且,实际物理系统中的温度涨落会抹平这些细微的差别,使得物理分布与 Benford 分布一致^[38]。因此,在满足三大物理统计的领域中,同样满足首位数分布定理。

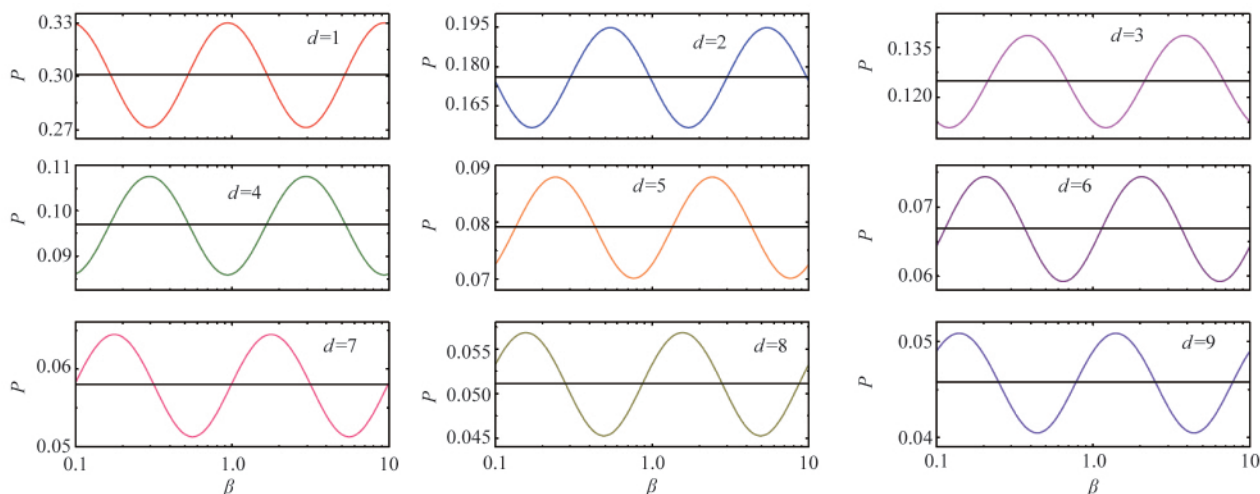


图 4 玻尔兹曼-吉布斯统计的首位数分布与 Benford 定理的比较

Fig. 4 Comparisons of the first digit distributions for Boltzmann-Gibbs distribution and Benford's law

6 结论与展望

与物理学的玻尔兹曼-吉布斯统计同时代诞生的 Benford 首位数定理,描述了自然界中绝大部分数据的首位数分布,

揭示了首位数分布不均匀的对数分布性质,被应用于社会学与自然科学的实际工作中。

首位数定理的提出至今已有百余年,但对它的理解尚未



完全,仍有很多问题需要进一步回答:为何首位数定理在大部分的自然现象中成立,而在小部分现象中不成立;如何先验地得知某个数据库是否符合 Benford 分布;为何指数尾数分布是负一次幂的分布;为何又存在推广的幂次分布,其幂指数 α 的物理含义是什么,它们与 Benford 数据库的本质差异在哪里等。对于这些问题的回答,有待于从事于物理学、数学、经济学、生物学等领域的研究人员的共同合作与努力。

参考文献 (References)

- [1] Benford F. The law of anomalous numbers [J]. *Proceedings of the American Philosophical Society*, 1938, 78(4): 551–572.
- [2] Torres J, Fernández S, Gamero A, et al. How do numbers begin? (The first digit law)[J]. *European Journal of Physics*, 2007, 28(3): L17–L25.
- [3] Newcomb S. Note on the frequency of use of the different digits in natural numbers[J]. *American Journal of Mathematics*, 1881, 4(1): 39–40.
- [4] Burke J, Kincanon E. Benford law and physical constants: The distribution of initial digits[J]. *American Journal of Physics*, 1991, 59: 952.
- [5] Buck B, Merchant A C, Perez S M. An illustration of Benford's first digit law using alpha decay half lives [J]. *European Journal of Physics*, 1993, 14: 59–63.
- [6] Ni D, Ren Z. Benford's law and half-lives of unstable nuclei [J]. *The European Physical Journal A*, 2008, 38: 251–255.
- [7] Ni D, Wei L, Ren Z. Benford's law and beta-decay half-lives[J]. *Communications in Theoretical Physics*, 2009, 51: 713–716.
- [8] Leemis L M, Schmeiser B W, Evans D L. Survival distributions satisfying Benford's law[J]. *The American Statistician*, 2000, 54: 236–241.
- [9] Pain J-C. Benford's law and complex atomic spectra [J]. *Physical Review E*, 2008, 77: 012102.
- [10] Giles D. Benford's law and naturally occurring prices in certain eBay auctions[J]. *Applied Economics Letters*, 2007, 14: 157–161.
- [11] Ley E. On the peculiar distribution of the US stock indexes' digits[J]. *The American Statistician*, 1996, 50: 311–313.
- [12] Tolle C R, Budzien J L, LaViolette R A. Do dynamical systems follow Benford's law?[J]. *Chaos*, 2000, 10: 331–336.
- [13] Berger A, Bunimovich L A, Hill T P. One-dimensional dynamical systems and Benford's law [J]. *Transactions of the American Mathematical Society*, 2005, 357: 197–219.
- [14] Berger A. Multi-dimensional dynamical systems and Benford's law[J]. *Discrete and Continuous Dynamical Systems*, 2005, 13: 219–237.
- [15] Nigrini M J. A taxpayer compliance application of Benford's law[J]. *The Journal of the American Taxation Association*, 1996, 18: 72–91.
- [16] Nigrini M J, Mittermaier L J. The use of Benford's law as an aid in analytical procedures[J]. *Auditing*, 1997, 16: 52–67.
- [17] Rose A M, Rose J M. Turn excel into a financial sleuth: An easy-to-use digital analysis tool can red-flag irregularities [J]. *Journal of Accountancy*, 2003, 196: 58–60.
- [18] Diekmann A. Not the first digit! Using Benford's law to detect fraudulent scientific data [J]. *Journal of Applied Statistics*, 2007, 34: 321–329.
- [19] Barlow J L, Bareiss E H. On roundoff error distributions in floating point and logarithmic arithmetic[J]. *Computing*, 1985, 34: 325–347.
- [20] Schatte P. On mantissa distributions in computing and Benford's law[J]. *Journal of Information Processing and Cybernetics*, 1988, 24 (9): 443–455.
- [21] Jolion J M. Images and Benford's law [J]. *Journal of Mathematical Imaging and Vision*, 2001, 14: 73–81.
- [22] Fu D, Shi Y Q, Su W. A generalized Benford's law for JPEG coefficients and its applications in image forensics [J]. *Proceedings of SPIE*, 2007, 6505: 65051L.
- [23] Amsler C, Doser M, Antonelli M, et al. Review of particle physics[J]. *Physics Letters B*, 2008, 667: 1–1340.
- [24] Shao L, Ma B Q. First digit distribution of hadron full width[J]. *Modern Physics Letters A*, 2009, 24(40): 3275–3282.
- [25] Manchester R N, Hobbs G B, Teoh A, et al. The australia telescope national facility pulsar catalogue [J]. *The Astronomical Journal*, 2005, 129(4): 1993–2006.
- [26] Shao L, Ma B Q. Empirical mantissa distributions of pulsars [J]. *Astroparticle Physics*, 2010, 33: 255–262.
- [27] Pinkham R S. On the distribution of first significant digits [J]. *The Annals of Mathematical Statistics*, 1961, 32: 1223–1230.
- [28] Berger A, Hill T P, Morrison K E. Scale-distortion inequalities for mantissas of finite data sets[J]. *Journal of Theoretical Probability*, 2008, 21: 97–117.
- [29] Raimi R A. On distribution of first significant figures [J]. *American Mathematical Monthly*, 1969, 76: 342–348.
- [30] Raimi R A. The first digit problem[J]. *American Mathematical Monthly*, 1976, 83: 521–538.
- [31] Raimi R A. The first digit phenomenon again [J]. *Proceedings of the American Philosophical Society*, 1985, 129: 211–219.
- [32] Hill T P. The significant-digit phenomenon [J]. *American Mathematical Monthly*, 1995, 102: 322–327.
- [33] Hill T P. Base-invariance implies Benford's law [J]. *Proceedings of the American Mathematical Society*, 1995, 123: 887–895.
- [34] Hill T P. A statistical derivation of the significant-digit law [J]. *Statistical Science*, 1995, 10: 354–363.
- [35] Hill T P. The first-digit phenomenon [J]. *American Scientist*, 1998, 86: 358.
- [36] Hill T P, Schurger K. Regularity of digits and significant digits of random variables[J]. *Stochastic Processes and Their Applications*, 2007, 115: 1723–1743.
- [37] Pietronero L, Tosatti E, Tosatti V, et al. Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf [J]. *Physica A*, 2001, 293(1–2): 297–304.
- [38] Shao L, Ma B Q. The significant digit law in statistical physics[J]. *Physica A*, 2010, 389(16): 3109–3116.
- [39] Engel H A, Leuenberger C. Benford's law for exponential random variables[J]. *Statistics & Probability Letters*, 2003, 63: 361–365.

(责任编辑 代丽)

《科技导报》“综述文章”栏目征稿

“综述文章”栏目发表对当前自然科学有关学科领域的研究热点、前沿分支发展现状及动向的评述性文章。要求在研读相当数量文献资料的基础上,全面、深入、系统地论述某一方面的问题,并对所综述的内容进行归纳、分析、评价,以反映作者的观点和见解。本栏目欢迎广大一线科技工作者投稿。投稿网址:www.kjdb.org;投稿信箱:kjdbbjb@cast.org.cn。